

RESCALED PURE GREEDY ALGORITHM FOR HILBERT AND BANACH SPACES

GUERGANA PETROVA

ABSTRACT. We show that a very simple modification of the Pure Greedy Algorithm for approximating functions by sparse sums from a dictionary in a Hilbert or more generally a Banach space has optimal convergence rates on the class of convex combinations of dictionary elements.

AMS subject classification: 41A25, 41A46.

Key Words: Greedy Algorithms, Rates of Convergence.

1. INTRODUCTION

Greedy algorithms have been used quite extensively as a tool for generating approximations from redundant families of functions, such as frames or more general dictionaries \mathcal{D} . Given a Banach space X , a dictionary is any set \mathcal{D} of norm one elements from X whose span is dense in X . The most natural greedy algorithm in a Hilbert space is the Pure Greedy Algorithm (**PGA**), which is also known as Matching Pursuit, see [2] for the description of this and other algorithms. The fact that the **PGA** lacks optimal convergence properties has led to a variety of modified greedy algorithms such as the Relaxed Greedy Algorithm (**RGA**), the Orthogonal Greedy Algorithm, and their weak versions. There are also analogues of these, developed for approximating functions in Banach spaces, see [10].

The central issues in the study of these algorithms is their ease of implementation and their approximation power, measured in terms of convergence rates. If f_m is the output of a greedy algorithm after m iterations, then f_m is a linear combination of at most m dictionary elements. Such linear combinations are said to be sparse of order m . The quality of the approximation is measured by the decay of the error $\|f - f_m\|$ as $m \rightarrow \infty$, where $\|\cdot\|$ is the norm in the Hilbert or Banach space, respectively. Of course, the decay rate of this error is governed by properties of the target function f . The typical properties imposed on f are that it is sparse, or more generally, that it is in some way compressible. Here, compressible means that it can be written as a (generally speaking, infinite) linear combination of dictionary elements with some restrictions on the coefficients. The most frequently applied assumption on f is that it is in the unit ball of the class $\mathcal{A}_1(\mathcal{D})$, that is the set of all functions which are a convex combination of dictionary elements (provided we consider symmetric dictionaries). It is known that the elements in this class can be approximated by m sparse vectors to accuracy $\mathcal{O}(m^{-1/2})$, see Theorem 2.1, and so this rate of approximation serves as a benchmark for the performance of greedy algorithms.

It has been shown in [2] in the case of Hilbert space that whenever $f \in \mathcal{A}_1(\mathcal{D})$, the output f_m of the **PGA** satisfies

$$(1.1) \quad \|f - f_m\| = \mathcal{O}(m^{-1/6}), \quad m \rightarrow \infty.$$

Later results gave slight improvements of the above estimate. For example, in [5], the rate $\mathcal{O}(m^{-1/6})$ was improved to $\mathcal{O}(m^{-11/62})$. Based on the method from the latter paper, Sil'nicenko [9] then

This research was supported by the Office of Naval Research Contract ONR N00014-11-1-0712, and by the NSF Grant DMS 1222715.

showed a rate of $\mathcal{O}(m^{-\frac{s}{2(s+2)}})$, where s solves a certain equation, and that $\frac{s}{2(s+2)} > 11/62$. Similar estimates for the weak versions of the **PGA** can be found in [10]. Estimates for the error from below have also been provided, see [7, 6].

The fact that the **PGA** does not attain the optimal rate for approximating the elements in $\mathcal{A}_1(\mathcal{D})$ has led to various modifications of this algorithm. Two of these modifications, the Relaxed and the Orthogonal Greedy Algorithm were shown to achieve the optimal rate $\mathcal{O}(m^{-1/2})$, see [2].

The purpose of the present paper is to show that a very simple modification of the **PGA**, namely just rescaling f_m at each iteration, already leads to the improved convergence rate $\mathcal{O}(m^{-1/2})$ for functions in $\mathcal{A}_1(\mathcal{D})$. The rescaling we suggest is simply the orthogonal projection of f onto f_m . We call this modified algorithm a Rescaled Pure Greedy Algorithm (**RPGA**) and prove optimal convergence rates for its weak version in Hilbert and Banach spaces. In a subsequent paper, see [4], we show that this strategy can also be applied successfully for developing an algorithm for convex optimization.

The paper is organized as follows. In §2, we spell out our notation and recall some simple known facts related to greedy algorithms. In §3, we present the **RPGA** for a Hilbert space and prove the above convergence rate. The remaining parts of this paper consider a modification of this algorithm for Banach spaces and weak versions of this algorithm.

2. NOTATION AND PRELIMINARIES

We denote by H a Hilbert space and by X a Banach space with $\|\cdot\|$ being the norm in these spaces, respectively. A set of functions $\mathcal{D} \subset H$ (or X) is called a dictionary if $\|\varphi\| = 1$ for every $\varphi \in \mathcal{D}$ and the closure of $\text{span}(\mathcal{D})$ is H (or X). An example of a dictionary is any Shauder basis for H (or X). However, the main idea behind dictionaries is to cover redundant families such as frames. A common example of dictionaries is the union of several Shauder bases.

The set $\Sigma_m(\mathcal{D})$ consists of all m -sparse elements with respect to the dictionary \mathcal{D} , namely

$$\Sigma_m := \Sigma_m(\mathcal{D}) = \{g : g = \sum_{\varphi \in \Lambda} c_\varphi \varphi, \Lambda \in \mathcal{D}, |\Lambda| \leq m\}.$$

Here, we use the notation $|\Lambda|$ to denote the cardinality of the index set Λ . For a general element f from X , we define the error of approximation

$$\sigma_m(f) := \sigma_m(f, \mathcal{D}) := \inf_{g \in \Sigma_m} \|f - g\|$$

of f by elements from Σ_m . The rate of decay of $\sigma_m(f)$ as $m \rightarrow \infty$ says how well f can be approximated by sparse elements.

For a general dictionary $\mathcal{D} \subset H$ (or X), we define the class of functions

$$\mathcal{A}_1^o(\mathcal{D}, M) := \{f = \sum_{k \in \Lambda} c_k(f) \varphi_k : \varphi_k \in \mathcal{D}, |\Lambda| < \infty, \sum_{k \in \Lambda} |c_k(f)| \leq M\},$$

and by $\mathcal{A}_1(\mathcal{D}, M)$ its closure in H (or X). Then, $\mathcal{A}_1(\mathcal{D})$ is defined to be the union of the classes $\mathcal{A}_1(\mathcal{D}, M)$ over all $M > 0$. For $f \in \mathcal{A}_1(\mathcal{D})$, we define the “semi-norm” of f as

$$|f|_{\mathcal{A}_1(\mathcal{D})} := \inf\{M : f \in \mathcal{A}_1(\mathcal{D}, M)\}.$$

A fundamental result for approximating $\mathcal{A}_1(\mathcal{D})$ is the following, see [2].

Theorem 2.1. *For a general dictionary $\mathcal{D} \subset H$ and $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, we have*

$$\sigma_m(f, \mathcal{D}) \leq c|f|_{\mathcal{A}_1(\mathcal{D})} m^{-1/2}, \quad m = 1, 2, \dots$$

When analyzing the convergence of greedy algorithms, we will use the following lemma, proved in [8].

Lemma 2.2. Let $\ell > 0$, $r > 0$, $B > 0$, and $\{a_m\}_{m=1}^\infty$ and $\{r_m\}_{m=2}^\infty$ be sequences of non-negative numbers satisfying the inequalities

$$a_1 \leq B, \quad a_{m+1} \leq a_m \left(1 - \frac{r_{m+1}}{r} a_m^\ell\right), \quad m = 1, 2, \dots$$

Then, we have

$$(2.1) \quad a_m \leq \max\{1, \ell^{-1/\ell}\} r^{1/\ell} (rB^{-\ell} + \sum_{k=2}^m r_k)^{-1/\ell}, \quad m = 2, 3, \dots$$

We note that several similar versions of this lemma have been proved and used in analysis of greedy algorithms, see [10].

3. THE HILBERT SPACE CASE

In order to show the simplicity of our results, we begin with the standard case of the **RPGA** in a Hilbert space. Later, we treat the case of Banach spaces and weak algorithms, but the reader familiar with this topic will see that the results in these more general settings follow by standard modifications of the results from this section. We denote the inner product in the Hilbert space H by $\langle \cdot, \cdot \rangle$, and so the norm of $f \in H$ is $\|f\| = \langle f, f \rangle^{1/2}$.

The **RPGA**(\mathcal{D}) is defined by the following simple steps.

RPGA(\mathcal{D}):

- **Step 0:** Define $f_0 := 0$.
- **Step m :**
- Assuming f_{m-1} has been computed and $f_{m-1} \neq f$. Choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

With

$$\lambda_m := \langle f - f_{m-1}, \varphi_m \rangle, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m, \quad s_m := \frac{\langle f, \hat{f}_m \rangle}{\|\hat{f}_m\|^2},$$

define the next approximant to be

$$f_m = s_m \hat{f}_m.$$

- If $f = f_m$, stop the algorithm and define $f_k = f_m = f$, for $k > m$.
- If $f \neq f_m$, proceed to Step $m + 1$.

Note that if the output at each Step m were \hat{f}_m and not $f_m = s_m \hat{f}_m$, this would be the **PGA**. However, the new algorithm uses not \hat{f}_m , but the best approximation to f from the one dimensional space $\text{span}\{\hat{f}_m\}$, that is $s_m \hat{f}_m$. Adding this step, which is just appropriate scaling of the output of the **PGA**, allows us to prove optimal convergence rate of $m^{-1/2}$ for the proposed algorithm.

Next, we show that the **RPGA** and the Relaxed Greedy Algorithm (**RGA**) provide different sequences of approximants $\{f_m\}$ and $\{f_m^r\}$, respectively, and thus **RPGA** is different from the known so far greedy algorithms. For both algorithms

$$f_0 = f_0^r = 0, \quad f_1 = f_1^r = \langle f, \varphi_1 \rangle \varphi_1,$$

where $\varphi_1 \in \mathcal{D}$ is such that $|\langle f, \varphi_1 \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f, \varphi \rangle|$. For both **RPGA** and **RGA**, the next element $\varphi_2 \in \mathcal{D}$ is chosen as $|\langle f - f_1, \varphi_2 \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_1, \varphi \rangle|$. One can easily compute that the next approximant, generated by the **RPGA** is

$$f_2 = s_2 f_1 + s_2 \langle f - f_1, \varphi_2 \rangle \varphi_2, \quad s_2 = \frac{\langle f, \varphi_1 \rangle^2 + \langle f, \varphi_2 \rangle^2 - \langle f, \varphi_1 \rangle \langle f, \varphi_2 \rangle \langle \varphi_1, \varphi_2 \rangle}{\langle f, \varphi_1 \rangle^2 + \langle f, \varphi_2 \rangle^2 - \langle f, \varphi_1 \rangle^2 \langle \varphi_1, \varphi_2 \rangle^2},$$

while the classical **RGA** would give

$$f_2^r = \frac{1}{2}f_1 + \frac{1}{2}\varphi_2.$$

There are some modifications of the **RGA**, see [1], where the approximant at Step m is determined not as

$$f_m^r = (1 - \frac{1}{m})f_{m-1}^r + \frac{1}{m}\varphi_m, \quad \text{where } |\langle f - f_{m-1}^r, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}^r, \varphi \rangle|,$$

but as

$$(3.1) \quad f_m^r = (1 - a_m)f_{m-1}^r + a_m\varphi_m,$$

where a_m and φ_m are the solutions of the minimization problem

$$\min_{a \in [0,1], \varphi \in \mathcal{D}} \|f - ((1-a)f_{m-1}^r + a\varphi)\|.$$

While the sequence, generated by the **RPGA** is a linear combination of f_{m-1} and φ_m , that is

$$f_m = s_m f_{m-1} + \lambda_m s_m \varphi_m,$$

it is different from the convex combinations (3.1), from other variations of the **RGA**, as described in [10], and from the best approximation to f from $\text{span}\{f_{m-1}^r, \varphi_m\}$. For example, the best approximation to f from $\text{span}\{f_1^r, \varphi_2\}$ is

$$f_2^r = \frac{\langle f, \varphi_1 \rangle^2 - \langle f, \varphi_1 \rangle \langle f, \varphi_2 \rangle \langle \varphi_1, \varphi_2 \rangle}{\langle f, \varphi_1 \rangle^2 (1 - \langle \varphi_1, \varphi_2 \rangle^2)} f_1 + \frac{\langle f, \varphi_2 \rangle - \langle f, \varphi_1 \rangle \langle \varphi_1, \varphi_2 \rangle}{1 - \langle \varphi_1, \varphi_2 \rangle^2} \varphi_2,$$

and again $f_2 \neq f_2^r$. In summary, we can view the new algorithm either as a rescaled version of the **PGA** or a new modification of the **RGA**.

We continue with the following theorem.

Theorem 3.1. *If $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, then the output $(f_m)_{m \geq 0}$ of the **RPGA**(\mathcal{D}) satisfies*

$$(3.2) \quad e_m := \|f - f_m\| \leq |f|_{\mathcal{A}_1(\mathcal{D})} m^{-1/2}, \quad m = 1, 2, \dots$$

Proof: Since f_m is the orthogonal projection of f onto the one dimensional space spanned by \hat{f}_m , we have

$$(3.3) \quad \langle f - f_m, f_m \rangle = 0, \quad m \geq 0.$$

Next, note that the definition of \hat{f}_m and the choice of λ_m give

$$\begin{aligned} \|f - \hat{f}_m\|^2 &= \langle f - f_{m-1} - \lambda_m \varphi_m, f - f_{m-1} - \lambda_m \varphi_m \rangle \\ &= \|f - \hat{f}_{m-1}\|^2 - 2\lambda_m \langle f - f_{m-1}, \varphi_m \rangle + \lambda_m^2 \|\varphi_m\|^2 \\ (3.4) \quad &= \|f - f_{m-1}\|^2 - \langle f - f_{m-1}, \varphi_m \rangle^2, \end{aligned}$$

where we have used that $\|\varphi_m\| = 1$. Now, assume $f \neq f_{m-1}$. Since f_m is the orthogonal projection of f onto $\text{span}\{\hat{f}_m\}$, we have

$$e_m^2 = \|f - f_m\|^2 = \|f - s_m \hat{f}_m\|^2 \leq \|f - \hat{f}_m\|^2.$$

We combine the latter inequality and (3.4) to derive that

$$(3.5) \quad e_m^2 \leq e_{m-1}^2 - \langle f - f_{m-1}, \varphi_m \rangle^2, \quad m = 1, 2, \dots$$

We proceed with an estimate from below for $\langle f - f_{m-1}, \varphi_m \rangle$. Note that

$$(3.6) \quad e_{m-1}^2 = \|f - f_{m-1}\|^2 = \langle f - f_{m-1}, f - f_{m-1} \rangle = \langle f - f_{m-1}, f \rangle,$$

where we have used (3.3).

It is enough to prove (3.2) for functions f that are finite sums $f = \sum_j c_j \varphi_j$ with $\sum_j |c_j| \leq M$, since these functions are dense in $\mathcal{A}_1(\mathcal{D}, M)$. Let us fix $\varepsilon > 0$ and choose a representation for $f = \sum_{\varphi \in \mathcal{D}} c_\varphi \varphi$, such that

$$\sum_{\varphi \in \mathcal{D}} |c_\varphi| < M + \varepsilon.$$

It follows from (3.6) that

$$\begin{aligned} e_{m-1}^2 &= \sum_{\varphi \in \mathcal{D}} c_\varphi \langle f - f_{m-1}, \varphi \rangle \\ &\leq |\langle f - f_{m-1}, \varphi_m \rangle| \sum_{\varphi \in \mathcal{D}} |c_\varphi| \\ &< |\langle f - f_{m-1}, \varphi_m \rangle| (M + \varepsilon), \end{aligned}$$

where we have used the choice of φ_m . We let $\varepsilon \rightarrow 0$ and obtain the inequality

$$(3.7) \quad M^{-1} e_{m-1}^2 \leq |\langle f - f_{m-1}, \varphi_m \rangle|.$$

We combine (3.5) and (3.7) to obtain

$$e_m^2 \leq e_{m-1}^2 - M^{-2} e_{m-1}^4 = e_{m-1}^2 (1 - M^{-2} e_{m-1}^2), \quad m \geq 2.$$

Note that

$$\|f\|^2 = \langle f, f \rangle = \sum_{\varphi \in \mathcal{D}} c_\varphi \langle f, \varphi \rangle \leq |\langle f, \varphi_1 \rangle| \sum_{\varphi \in \mathcal{D}} |c_\varphi| < \|f\| (M + \varepsilon),$$

and therefore $\|f\| \leq M$. Since $e_1^2 \leq e_0^2 = \|f\|^2 \leq M^2$, we can apply Lemma 2.2 with $a_m = e_m^2$, $B = M^2$, $r_m := 1$, $r = M^2$, and $\ell = 1$. Then, (2.1) gives

$$e_m^2 \leq M^2 m^{-1}, \quad m \geq 2,$$

and the theorem follows. \square

In the sections that follow, we introduce variants of the **RPGA** and prove convergence results similar to Theorem 3.1.

4. THE WEAK RESCALED PURE GREEDY ALGORITHM FOR HILBERT SPACES

In this section, we describe the Weak Rescaled Pure Greedy Algorithm (**WRPGA**). It is determined by a weakness sequence $\{t_k\}_{k=1}^\infty$, where all $t_k \in (0, 1]$, and the dictionary \mathcal{D} . We denote it by **WRPGA**($\{t_k\}, \mathcal{D}$).

WRPGA($\{t_k\}, \mathcal{D}$):

- **Step 0:** Define $f_0 = 0$.
- **Step m :**
- Assuming f_{m-1} has been computed and $f_{m-1} \neq f$. Choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|\langle f - f_{m-1}, \varphi_m \rangle| \geq t_m \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}, \varphi \rangle|.$$

With

$$\lambda_m = \langle f - f_{m-1}, \varphi_m \rangle, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m, \quad s_m = \frac{\langle f, \hat{f}_m \rangle}{\|\hat{f}_m\|^2},$$

define the next approximant to be

$$f_m = s_m \hat{f}_m.$$

- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
- If $f \neq f_m$, proceed to Step $m + 1$.

In the case when all elements t_k of the weakness sequence are $t_k = 1$, this algorithm is the **RPGA**(\mathcal{D}). The following theorem holds.

Theorem 4.1. *If $f \in \mathcal{A}_1(\mathcal{D}) \subset H$, then the output $(f_m)_{m \geq 0}$ of the **WRPGA**($\{t_k\}, \mathcal{D}$) satisfies*

$$(4.1) \quad e_m := \|f - f_m\| \leq |f|_{\mathcal{A}_1(\mathcal{D})} \left(\sum_{k=1}^m t_k^2 \right)^{-1/2}, \quad m \geq 1.$$

Proof: The proof is similar to the one of Theorem 3.1, where we show that for the error $e_m^2 = \|f - f_m\|^2$, we have the inequality,

$$(4.2) \quad e_m^2 \leq e_{m-1}^2 - \langle f - f_{m-1}, \varphi_m \rangle^2, \quad m = 1, 2, \dots$$

The estimate from below for $\langle f - f_{m-1}, \varphi_m \rangle$ is derived similarly as

$$(4.3) \quad M^{-1} t_m e_{m-1}^2 \leq |\langle f - f_{m-1}, \varphi_m \rangle|,$$

where we have used the definition of φ_m . Next, it follows from (4.2) and (4.3) that

$$e_m^2 \leq e_{m-1}^2 - M^{-2} t_m^2 e_{m-1}^4 = e_{m-1}^2 (1 - M^{-2} t_m^2 e_{m-1}^2), \quad m \geq 1,$$

Note that

$$\|f\|^2 = \langle f, f \rangle = \sum_{\varphi \in \mathcal{D}} c_\varphi \langle f, \varphi \rangle \leq t_1^{-1} |\langle f, \varphi_1 \rangle| \sum_{\varphi \in \mathcal{D}} |c_\varphi| < t_1^{-1} \|f\| (M + \varepsilon),$$

and therefore $\|f\| \leq M t_1^{-1}$. Since $e_1^2 \leq e_0^2 = \|f\|^2 \leq M^2 t_1^{-2}$, we can apply Lemma 2.2 with $a_m = e_m^2$, $B = M^2 t_1^{-2}$, $r_m := t_m^2$, $r = M^2$, and $\ell = 1$ to obtain

$$e_m^2 \leq M^2 \left(t_1^2 + \sum_{k=2}^m t_k^2 \right)^{-1}, \quad m \geq 2,$$

and the theorem follows. \square

5. THE BANACH SPACE CASE

In this section, we will state the **RPGA**(\mathcal{D}) algorithm for Banach spaces X with norm $\|\cdot\|$ and dictionary \mathcal{D} , and prove convergence results for certain Banach spaces. Let us first start with the introduction of the modulus of smoothness ρ of a Banach space X , which is defined as

$$\rho(u) := \sup_{f, g \in X, \|f\| = \|g\| = 1} \left\{ \frac{1}{2} (\|f + ug\| + \|f - ug\|) - 1 \right\}, \quad u > 0.$$

In this paper, we shall consider only Banach spaces X whose modulus of smoothness satisfies the inequality

$$\rho(u) \leq \gamma u^q, \quad 1 < q \leq 2, \quad \gamma \text{ -constant.}$$

This is a natural assumption, since the modulus of smoothness of $X = L_p$, $1 < p < \infty$, for example, is known to satisfy such inequality. Recall that, see [3], for $X = L_p$,

$$\rho(u) \leq \begin{cases} \frac{1}{p} u^p, & \text{if } 1 \leq p \leq 2, \\ \frac{p-1}{2} u^2, & \text{if } 2 \leq p < \infty. \end{cases}$$

Next, for every element $f \in X$, $f \neq 0$, we consider its norming functional $F_f \in X^*$ with the properties $\|F_f\| = 1$, $F_f(f) = \|f\|$. Note that if $X = H$ is a Hilbert space, the norming functional for $f \in H$ is

$$F_f(\cdot) = \frac{\langle f, \cdot \rangle}{\|f\|}.$$

There is a relationship between the norming functional F_g for any $g \in X$, $g \neq 0$, and the modulus of smoothness of X , given by the following lemma.

Lemma 5.1. *Let X be a Banach space with modulus of smoothness ρ , where $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Let $g \in X$, $g \neq 0$ with norming functional F_g . Then, for every $h \in X$, we have*

$$(5.1) \quad \|g + uh\| \leq \|g\| + uF_g(h) + 2\gamma u^q \|g\|^{1-q} \|h\|^q, \quad u > 0.$$

Proof: The proof follows from Lemma 6.1 in [10] and the property of the modulus of smoothness. \square

We next present the **RPGA**(\mathcal{D}) for the Banach space X with dictionary \mathcal{D} .

RPGA(\mathcal{D}):

- **Step 0:** Define $f_0 = 0$.
- **Step m :**
- Assuming f_{m-1} has been computed and $f \neq f_{m-1}$. Choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|F_{f-f_{m-1}}(\varphi_m)| = \sup_{\varphi \in \mathcal{D}} |F_{f-f_{m-1}}(\varphi)|.$$

With

$$\lambda_m = \text{sign}\{F_{f-f_{m-1}}(\varphi_m)\} \|f - f_{m-1}\| (2\gamma q)^{\frac{1}{1-q}} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{1}{q-1}}, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m,$$

choose s_m such that

$$\|f - s_m \hat{f}_m\| = \min_{s \in \mathbb{R}} \|f - s \hat{f}_m\|,$$

and define the next approximant to be

$$f_m = s_m \hat{f}_m.$$

- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
- If $f \neq f_m$, proceed to Step $m + 1$.

The following lemma holds.

Lemma 5.2. *Let X be a Banach space with modulus of smoothness ρ , $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. Let f_{m-1} be the output of the **RPGA**(\mathcal{D}) at Step $m - 1$. Then, if $f \neq f_{m-1}$, we have*

$$F_{f-f_{m-1}}(f_{m-1}) = 0.$$

Proof: Let us denote by $L := \text{span}\{\hat{f}_{m-1}\} \subset X$. Clearly, $f_{m-1} \in L$, and moreover, f_{m-1} is the best approximation to f from L . We apply Lemma 6.9 from [10] to the linear space L and the vector f_{m-1} , and derive the lemma. \square

The next theorem provides the convergence rate for the new algorithm in Banach spaces.

Theorem 5.3. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. If $f \in \mathcal{A}_1(\mathcal{D}) \subset X$, then the output $(f_m)_{m \geq 0}$ of the **RPGA**(\mathcal{D}) satisfies*

$$(5.2) \quad e_m := \|f - f_m\| \leq c \|f\|_{\mathcal{A}_1(\mathcal{D})} m^{1/q-1}, \quad m \geq 2,$$

where $c = c(\gamma, q)$.

Proof: Clearly, we have $e_0 = \|f - f_0\| = \|f\|$. At Step m , $m = 1, 2, \dots$ of the algorithm, either $f = f_{m-1}$, in which case $f_k = f_{m-1}$, $k \geq m$, and therefore $e_m = 0$, or we have

$$e_m = \|f - f_m\| = \|f - s_m \hat{f}_m\| \leq \|f - \hat{f}_m\| = \|(f - f_{m-1}) - \lambda_m \varphi_m\|.$$

We now apply Lemma 5.1 to the latter inequality with $g = f - f_{m-1} \neq 0$, $u = |\lambda_m| > 0$, $h = -\text{sign}\{\lambda_m\}\varphi_m$, and derive

$$\begin{aligned}
e_m &\leq \|f - f_{m-1}\| - \lambda_m F_{f-f_{m-1}}(\varphi_m) + 2\gamma|\lambda_m|^q \|f - f_{m-1}\|^{1-q} \|\varphi_m\|^q \\
&= e_{m-1} - \lambda_m F_{f-f_{m-1}}(\varphi_m) + 2\gamma|\lambda_m|^q e_{m-1}^{1-q} \\
(5.3) \quad &= e_{m-1} - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} e_{m-1} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{q}{q-1}},
\end{aligned}$$

where we have used that $\|\varphi_m\| = 1$ and the choice of λ_m . Now, we need an estimate from below for $|F_{f-f_{m-1}}(\varphi_m)|$. Using Lemma 5.2, we obtain that

$$(5.4) \quad e_{m-1} = \|f - f_{m-1}\| = F_{f-f_{m-1}}(f - f_{m-1}) = F_{f-f_{m-1}}(f).$$

As in the Hilbert space case, it is enough to consider functions f that are finite sums $f = \sum_j c_j \varphi_j$ with $\sum_j |c_j| \leq M$, since these functions are dense in $\mathcal{A}_1(\mathcal{D}, M)$. Let us fix $\varepsilon > 0$ and choose a representation for $f = \sum_{\varphi \in \mathcal{D}} c_\varphi \varphi$, such that

$$\sum_{\varphi \in \mathcal{D}} |c_\varphi| < M + \varepsilon.$$

It follows that

$$\begin{aligned}
F_{f-f_{m-1}}(f) &= \sum_{\varphi \in \mathcal{D}} c_\varphi F_{f-f_{m-1}}(\varphi) \leq \sum_{\varphi \in \mathcal{D}} |c_\varphi| |F_{f-f_{m-1}}(\varphi)| \\
&\leq |F_{f-f_{m-1}}(\varphi_m)| \sum_{\varphi \in \mathcal{D}} |c_\varphi| < |F_{f-f_{m-1}}(\varphi_m)| (M + \varepsilon).
\end{aligned}$$

We take $\varepsilon \rightarrow 0$ and derive

$$F_{f-f_{m-1}}(f) \leq |F_{f-f_{m-1}}(\varphi_m)| M$$

The latter estimate and (5.4) provide the estimate from below

$$M^{-1} e_{m-1} \leq |F_{f-f_{m-1}}(\varphi_m)|,$$

which together with (5.3) result in

$$e_m \leq e_{m-1} \left(1 - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} M^{-\frac{q}{q-1}} e_{m-1}^{\frac{q}{q-1}} \right).$$

Note that $e_1 \leq e_0 = \|f\| \leq M$, since

$$\|f\| = F_f(f) = \sum_{\varphi} c_\varphi F_f(\varphi) \leq |F_f(\varphi_1)| \sum_{\varphi} |c_\varphi| < M + \varepsilon,$$

for every $\varepsilon > 0$. We now use Lemma 2.2 with $a_m = e_m$, $B = M$, $r_m := \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}}$, $r = M^{\frac{q}{q-1}}$, and $\ell = \frac{q}{q-1}$ to obtain

$$e_m \leq M \left(1 + \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} (m-1) \right)^{1/q-1}, \quad m \geq 2,$$

and the theorem follows. □

6. THE WEAK RESCALED PURE GREEDY ALGORITHM FOR BANACH SPACES

In this section, we describe the Weak Rescaled Pure Greedy Algorithm for Banach spaces. It is determined by a weakness sequence $\{t_k\}_{k=1}^\infty$, where all $t_k \in (0, 1]$, and the dictionary \mathcal{D} . As in the Hilbert case, we denote it by **WRPGA** $(\{t_k\}, \mathcal{D})$.

WRPGA $(\{t_k\}, \mathcal{D})$:

- **Step 0**: Define $f_0 = 0$.
- **Step m** :
- Assuming f_{m-1} has been computed and $f \neq f_{m-1}$. Choose a direction $\varphi_m \in \mathcal{D}$ such that

$$|F_{f-f_{m-1}}(\varphi_m)| \geq t_m \sup_{\varphi \in \mathcal{D}} |F_{f-f_{m-1}}(\varphi)|.$$

With

$$\lambda_m = \text{sign}\{F_{f-f_{m-1}}(\varphi_m)\} \|f - f_{m-1}\| (2\gamma q)^{\frac{1}{1-q}} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{1}{q-1}}, \quad \hat{f}_m := f_{m-1} + \lambda_m \varphi_m,$$

choose s_m such that

$$\|f - s_m \hat{f}_m\| = \min_{s \in \mathbb{R}} \|f - s \hat{f}_m\|,$$

and define the next approximant to be

$$f_m = s_m \hat{f}_m.$$

- If $f = f_{m-1}$, stop the algorithm and define $f_k = f_{m-1} = f$ for $k \geq m$.
- If $f \neq f_m$, proceed to Step $m + 1$.

Next, we present the convergence rates for the **WRPGA** $(\{t_k\}, \mathcal{D})$ in Banach Spaces.

Theorem 6.1. *Let X be a Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q$, $1 < q \leq 2$. If $f \in \mathcal{A}_1(\mathcal{D}) \subset X$, then the output $(f_m)_{m \geq 0}$ of the **WRPGA** $(\{t_k\}, \mathcal{D})$ satisfies*

$$(6.1) \quad e_m := \|f - f_m\| \leq c \|f\|_{\mathcal{A}_1(\mathcal{D})} \left(\sum_{k=1}^m t_k^{\frac{q}{q-1}} \right)^{1/q-1}, \quad m \geq 1,$$

where $c = c(\gamma, q)$.

Proof: As in the proof of Theorem 5.3, we show that

$$(6.2) \quad e_m \leq e_{m-1} - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} e_{m-1} |F_{f-f_{m-1}}(\varphi_m)|^{\frac{q}{q-1}}.$$

Next, similarly to Theorem 5.3, we prove an estimate from below for $|F_{f-f_{m-1}}(\varphi_m)|$, which is

$$M^{-1} t_m e_{m-1} \leq |F_{f-f_{m-1}}(\varphi_m)|,$$

which together with (6.2) result in

$$e_m \leq e_{m-1} \left(1 - \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} t_m^{\frac{q}{q-1}} M^{-\frac{q}{q-1}} e_{m-1}^{\frac{q}{q-1}} \right).$$

Again, since $e_1 \leq e_0 = \|f\| \leq M t_1^{-1}$, we can use Lemma 2.2 with $a_m = e_m$, $B = M t_1^{-1}$, $r_m := \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} t_m^{\frac{q}{q-1}}$, $r = M^{\frac{q}{q-1}}$, and $\ell = \frac{q}{q-1}$. Then, (2.1) gives

$$e_m \leq M \left(t_1^{\frac{q}{q-1}} + \frac{q-1}{q} (2\gamma q)^{\frac{1}{1-q}} \sum_{k=2}^m t_k^{\frac{q}{q-1}} \right)^{1/q-1}, \quad m \geq 2,$$

and the theorem follows. □

REFERENCES

- [1] A. Barron, *Universal approximation bounds for superposition of n sigmoidal functions*, IEEE Trans. Inf. Theory, **39**, 930-945, 1993.
- [2] R. DeVore, V. Temlyakov, *Some remarks on greedy algorithms*, Advances in Computational Math., **5**, 173-187, 1996.
- [3] M. Donahue, L. Gurvits, C. Darken, E. Sontag, *Rate of convex approximation in non-Hilbert spaces*, Constr. Approx., **13**, 187-220, 1997.
- [4] Z. GAO, G. Petrova, *Rescaled Pure Greedy Algorithm for Convex Optimization*, submitted.
- [5] S. Konyagin, V. Temlyakov, *Rate of convergence of Pure Greedy Algorithm*, East. J. Approx., **5**, 493-499, 1999.
- [6] E. Livshitz, *On lower estimates of rate of convergence of greedy algorithms*, Izv. RAN, Ser. Matem, **73**, 125-144, 2009.
- [7] E. Livshitz, V. Temlyakov, *Two lower estimates in greedy approximation*, Constr. Approx. **19**, 509-523, 2003.
- [8] H. Nguyen, G. Petrova, *Greedy strategies for convex optimization*, submitted.
- [9] Sil'nichenko, *Rates of convergence of greedy algorithms*, Mat. Zametki, **76**, 628-632, 2004.
- [10] V. Temlyakov, *Greedy approximation*, Cambridge monographs on Applied and Computational Mathematics, Cambridge University Press, 2011.

Guergana Petrova

Department of Mathematics, Texas A&M University, College Station, TX 77843, USA
 gpetrova@math.tamu.edu